

Noncovalent Interactions by QMC: Speedup by One-Particle Basis-Set Size Reduction

Matúš Dubecký*

*Department of Physics, Faculty of Science, University of Ostrava,
30. dubna 22, 701 03 Ostrava, Czech Republic and*

*RCPTM, Department of Physical Chemistry, Faculty of Science,
Palacký University Olomouc, tř. 17 listopadu 12, 771 46 Olomouc, Czech Republic*

(Dated: September 12, 2016)

While it is empirically accepted that the fixed-node diffusion Monte-Carlo (FN-DMC) depends only weakly on the size (beyond a certain reasonable level) of the one-particle basis sets used to expand its guiding functions, limits of this observation are not settled yet. Our recent work indicates that under the FN error cancellation conditions, augmented triple zeta basis sets are sufficient to achieve high-quality benchmark single-point energy differences in a number of small noncovalent complexes. In this preliminary progress report, we report on a possibility of significant truncation of the one-particle basis sets used to express the FN-DMC guiding functions, that has no visible effect on the accuracy of the production energy differences. The proposed scheme shows negligible increase of the local energy variance, indicating that the total CPU cost of large-scale benchmark noncovalent interaction energy FN-DMC calculations employing Gaussians may be reduced.

In the domain of benchmark ab-initio noncovalent interaction energy calculations, fixed-node diffusion Monte Carlo (FN-DMC) method provides a promising alternative to the commonly used coupled-cluster (CC) approaches like CCSD(T)¹. Accurate FN-DMC interaction energies (to 0.1 kcal/mol vs. CCSD(T)/CBS) have been recently reported on a number of small/medium noncovalent closed-shell complexes²⁻⁹. In addition to accuracy, FN-DMC is attractive also for its favorable low-order polynomial CPU cost scaling¹⁰⁻¹² and favorable FN error cancellation¹³⁻¹⁵ that enabled its use in medium/large complexes^{3,8,9,16-18} where CC methods were intractable (in original formulation and reasonable basis set) until recently¹⁹. Furthermore, direct treatment of extended^{9,20-28} and/or multireference systems²⁹ makes FN-DMC an attractive many-body method worth of further research and development.

While it is empirically accepted that the FN-DMC results depend only weakly on the one-particle basis sets¹ used to expand the FN-DMC guiding functions, limits of this assumption in energy differences remain unclear. Our recent work e.g. indicates that in FN-DMC calculations using DFT-based single-determinant guiding functions, the basis set cardinality is not as important as the presence of augmentation functions¹⁵. An example of ammonia dimer complex well illustrates this behavior: a sequence of VTZ, VQZ and aug-VTZ basis sets generates the interaction energies of -3.33 ± 0.07 , -3.47 ± 0.07 and -3.10 ± 0.06 kcal/mol¹⁵, while the complete-basis-set(CBS) CCSD(T) benchmark value at the same geometry amounts to -3.15 kcal/mol^{30,31}. Augmented triple zeta basis sets were confirmed to be sufficient to achieve a level of 0.1 kcal/mol in the final FN-DMC interaction energies in numerous cases^{7,15}. Some residual errors that remain in certain types of noncovalent interactions (e.g. stacking or hydrogen bonds combined with multiple bonding) are not yet understood and require further attention³².

QMC is nevertheless a very promising methodology,

and its main limitation in area of noncovalent interactions is the CPU cost that stems from the tight statistical convergence required in case of small energy differences. It is therefore important to map out strategies of possible CPU cost reduction. For instance, presence of the FN error cancellation^{1,7,13,14,32} enabled use of economic Jastrow factor with electron-electron and electron-nucleus terms instead of the more demanding one including also electron-electron-nucleus terms¹⁵.

Here we report on a possibility of truncation of the one-particle Gaussian basis sets used to expand the orbitals in the single Slater determinant FN-DMC guiding functions without affecting the accuracy of the final energy differences. A series of tests⁷ led us to the finding that high angular momentum basis functions are not critically important^{14,33} and the augmentation basis set is necessary, but the acceptable accuracy is achieved already with a single diffuse *s* function per atom used instead of the common aug- set of basis functions⁷.

In the following, we compare the two types of basis sets used to express the orbitals in single-determinant FN-DMC guiding functions. The results obtained with trimmed basis set (e.g. [3s3p2d]+[1s] on carbon atom) denoted as s-rVTZ are compared to the ones obtained with the standard aug-VTZ basis set ([3s3p2d1f]+[1s1p1d1f] on carbon atom). Surprisingly, the (single-point) interaction energies produced with the trimmed bases reveal no statistically significant differences with respect to the reference calculations, and, no significant increase of the local energy variance, implying that the equivalent-quality FN-DMC results are available at costs lower than assumed to date.

The considered test set contains seven noncovalent closed-shell complexes (with geometries obtained) from the A24 database³⁴: ammonia dimer (AM...AM), water-ammonia (WA...AM), water dimer (WA...WA), ammonia methane complex (AM...MT), methane dimer (MT...MT), hydrogen fluoride dimer (HF...HF), and HCN dimer (HCN...HCN). The atomic cores in these

complexes were replaced by the effective core potentials (ECPs) developed by Burkatzki et al.^{35,37}. The ECPs were used in combination with Dunning-type aug-VTZ basis sets or their truncated counterparts (s-rVTZ, see above). Single-determinant Slater-Jastrow¹¹ trial wave functions were constructed using orbitals from B3LYP (GAMESS³⁶ code) and the Schmidt-Moskowitz³⁷ isotropic Jastrow factors¹¹ including electron-electron and electron-nucleus terms¹⁵ were expanded in a fixed basis set of polynomial Padé functions¹¹. The Jastrow parameters were refined by the Hessian driven variational Monte Carlo optimization using at least 10x10 iterations and linear combination³⁸ of energy (95%) and variance (5%) as a cost function. The orbital parameters were frozen to keep the nodes of the guiding functions intact. The production FN-DMC runs used time step of 0.005 a.u. and the T-moves scheme for the treatment of ECPs³⁹. The target walker populations in FN-DMC amounted to 16-32k. All QMC calculations were performed using the code QWalk⁴⁰.

The FN-DMC results obtained with single-determinant Slater-Jastrow guiding functions expanded in aug-VTZ and reduced s-rVTZ basis sets are reported in Table I. The interaction energies obtained with s-rVTZ basis are clearly compatible with the aug-VTZ data in *all* considered cases (Figure 1), that is a non-trivial and very interesting observation, since, in the trimmed case, the total number of basis functions is reduced by more than two times, and, the local energy variance increases only by up to 4% (Table I).

Why is this the case? As mentioned above, VTZ or even VQZ bases do not lead to the correct energy differences in the similar setting, so the non-augmented rVTZ basis sets could hardly lead to the more accurate results. We thus presumably attribute the observed agreement of s-rVTZ and aug-VTZ results to the similar degree of FN error cancellation and sufficient sampling of interstitial regions secured by the single *s* function per atom, working even in such a complex like HF dimer⁷. The modest increase of σ^2 is attributed to the use of identical occupied-shell contractions.

Let us discuss the CPU cost benefits of the proposed basis-set size trimming. The Slater matrix evaluation step scales as $O(MN^3)$ where M is the number of basis functions and N is the number of electrons. The ideal speedup with reduced M for asymptotically large fixed N is proportional to

$$s_i = \frac{\sigma_{\text{ref}}^2 M_{\text{ref}}}{\sigma_{\text{trim}}^2 M_{\text{trim}}} \quad (1)$$

where M_{ref} is the number of basis functions in the reference basis set (aug-VTZ), M_{trim} is the count of the basis functions in the trimmed case (s-rVTZ), and σ^2 s denote the respective local energy variances. As the formula (1) indicates, the cost savings start to become interesting for M_{trim} significantly smaller than M_{ref} and only if the ratio of local energy variances $\sigma_{\text{ref}}^2/\sigma_{\text{trim}}^2$ does not outweigh the gain from $M_{\text{ref}}/M_{\text{trim}}$. Furthermore, the

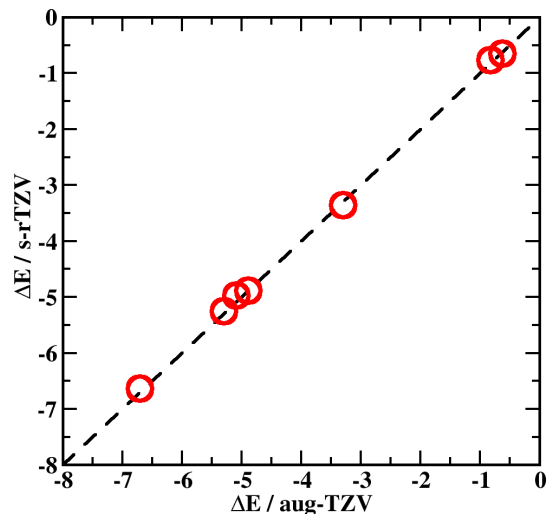


Figure 1. Demonstration of the agreement between the FN-DMC interaction energies ΔE (kcal/mol) from Table I. calculated with aug-VTZ and truncated s-rVTZ basis sets. The error bars (not shown) are smaller than the symbol size.

observed speedup is smaller than the theoretical limit s_i , since, other routines, like evaluation of Jastrow factor, pseudopotentials and/or distance matrix updates, also add to the overall cost. The observed speedup thus approaches s_i only for sufficiently large M where the Slater matrix updates dominate the overall CPU cost. Even in the theoretical limit, the CPU cost gain grows linearly that may appear unimportant. Nevertheless, the FN-DMC calculations are very expensive in general, and frequently millions of CPU-hours are invested in valuable projects (e.g. Ref. 18). Therefore, in our opinion, *any* CPU cost reduction is important, and the scheme presented here (i.e. use of s-rVTZ instead of aug-VTZ, in combination with Jastrow factor containing electron-electron and electron-nucleus terms) asymptotically offers about two-fold speedup (cf. Table I and examples below).

The following examples illustrate the wave function evaluation speedups (compare to $M_{\text{ref}}/M_{\text{trim}}$) in pure Slater and Slater-Jastrow runs:

i) In the HCN dimer (20 electrons) where $M_{\text{ref}}/M_{\text{trim}} = 2.02$, the practical Slater evaluations with s-rVTZ are 1.5 times faster than in the case of aug-VTZ and the Slater-Jastrow run achieves an evaluation cost speedup of 1.22, rather far from the expected value. This indicates that the system is “small” and routines other than the Slater matrix value and Laplacian updates are still important.

ii) In the case of a larger complex, coronene...H₂ with 110 electrons ($M_{\text{ref}}/M_{\text{trim}} = 1402/692 = 2.03$), the speedup of a pure Slater run is 1.92-fold while the FN-DMC calculation using Slater-Jastrow guiding function achieves a speedup factor of 1.74. Note that the local energy variance ratio $\sigma_{\text{ref}}^2/\sigma_{\text{trim}}^2 = 2.590/2.707 = 0.96$ so the total observed speedup of the calculation achieves a

Table I. Comparison of interaction energies ΔE (kcal/mol) and local energy variances in dimers σ^2 (a.u.²) from FN-DMC calculations using aug-VTZ (ΔE s taken from Ref. 15) and trimmed, s-rVTZ, basis sets, the related counts of the basis functions M in the dimer of the given complex, and, ideal expected speedup (s_i , see text) of s-rVTZ vs. aug-VTZ calculation.

Complex	aug-VTZ			s-rVTZ			s_i
	ΔE	σ^2	M	ΔE	σ^2	M	
AM...AM	-3.30±0.04	0.433	228	-3.36±0.08	0.445	106	2.09
AM...WA	-6.71±0.07	0.561	205	-6.64±0.09	0.578	96	2.07
WA...WA	-5.30±0.05	0.667	182	-5.25±0.09	0.682	86	2.07
HF...HF	-4.89±0.05	0.960	136	-4.89±0.10	0.968	66	2.04
AM...MT	-0.83±0.06	0.360	251	-0.77±0.07	0.364	106	2.34
MT...MT	-0.63±0.03	0.271	274	-0.65±0.04	0.282	126	2.09
HCN...HCN	-5.09±0.08	0.582	226	-4.97±0.08	0.605	112	1.94

factor of 1.67 ($s_i=1.94$). In larger systems, the speedups clearly grow toward the theoretical limit and become interesting.

In conclusion, a new one-particle basis set truncation scheme has been shown to considerably reduce computational requirements while retaining a full accuracy in energy differences of seven noncovalent complexes. This finding presumably extends the range of applicability for FN-DMC method in area of noncovalent interactions. Further tests including extended sets of molecules, other truncation schemes, and various combinations of tuned bases with various Jastrows are now underway[?].

ACKNOWLEDGMENTS

The author is grateful to Claudia Filippi for sharing an improved ECP for H and to Lubos Mitas, Petr Jurečka, and René Derian for stimulating and fruitful discussions. Financial support from the University of Ostrava (IRP201558), Ministry of Education, Youth and Sports of the Czech Republic (project LO1305), VEGA (VEGA-2/0130/15), and, computational resources provided by the Ministry of Education, Youth and Sports of the Czech Republic under the Projects CES-NET (LM2015042) and CERIT-Scientific Cloud (LM2015085) provided within the program Projects of Large Research, Development and Innovations Infrastructures, are gratefully acknowledged.

* matus.dubecky@osu.cz

- ¹ M. Dubecký, L. Mitas, and P. Jurečka. *Chem. Rev.*, 116:5188–5215, 2016.
- ² I. G. Gurtubay and R. J. Needs. *J. Chem. Phys.*, 127:124306, 2007.
- ³ B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, C. Filippi, and M. Scheffler. *J. Chem. Phys.*, 129:194111, 2008.
- ⁴ J. Ma, D. Alfè, A. Michaelides, and E. Wang. *J. Chem. Phys.*, 130:154303, 2009.
- ⁵ M. Korth, S. Grimme, and M. D. Towler. *J. Phys. Chem. A*, 115:11734–11739, 2011.
- ⁶ M. J. Gillan, F. R. Manby, M. D. Towler, and D. Alfè. *J. Chem. Phys.*, 136:244105, 2012.
- ⁷ M. Dubecký, P. Jurečka, R. Derian, P. Hobza, M. Otyepka, and L. Mitas. *J. Chem. Theory Comput.*, 9:4287–4292, 2013.
- ⁸ A. Benali, L. Shulenburger, N. A. Romero, J. Kim, and O. A. von Lilienfeld. *J. Chem. Theory Comput.*, 10:3417–3422, 2014.
- ⁹ Y. S. Al-Hamdani, D. Alfè, O. A. von Lilienfeld, and A. Michaelides. *J. Chem. Phys.*, 141:18C530, 2014.
- ¹⁰ W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. *Rev. Mod. Phys.*, 73:33–83, Jan 2001.
- ¹¹ M. Bajdich and L. Mitas. *Acta Phys. Slovaca*, 59:81–168, 2009.
- ¹² B. M. Austin, D. Y. Zubarev, and W. A. Lester. *Chem. Rev.*, 112:263–288, 2012.

- ¹³ M. Mella and J. B. Anderson. *J. Chem. Phys.*, 119:8225–8228, 2003.
- ¹⁴ C. Diedrich, A. Lühchow, and S. Grimme. *J. Chem. Phys.*, 123:184106, 2005.
- ¹⁵ M. Dubecký, R. Derian, P. Jurečka, P. Hobza, M. Otyepka, and L. Mitas. *Phys. Chem. Chem. Phys.*, 16:20915–20923, 2014.
- ¹⁶ S. Sorella, M. Casula, and D. Rocca. *J. Chem. Phys.*, 127:014105, 2007.
- ¹⁷ M. Korth, A. Lühchow, and S. Grimme. *J. Phys. Chem. A*, 112:2104–2109, 2008.
- ¹⁸ A. Ambrosetti, D. Alfè, R. A. DiStasio Jr., and A. Tkatchenko. *J. Phys. Chem. Lett.*, 5:849–855, 2014.
- ¹⁹ J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G.K.-L. Chan. *Science*, 345:640–643, 2014.
- ²⁰ Z. Raza, D. Alfè, C. G. Salzmann, J. Klimeš, A. Michaelides, and B. Slater. *Phys. Chem. Chem. Phys.*, 13:19788–19795, 2011.
- ²¹ B. Santra, J. Klimeš, D. Alfè, A. Tkatchenko, B. Slater, A. Michaelides, R. Car, and M. Scheffler. *Phys. Rev. Lett.*, 107:185701, 2011.
- ²² K. Hongo, M. A. Watson, R. S., T. Iitaka, A. Aspuru-Guzik, and R. Maezono. *J. Chem. Theory Comput.*, 11:907–917, 2015.
- ²³ O. Karalti, D. Alfè, M. J. Gillan, and K. D. Jordan. *Phys. Chem. Chem. Phys.*, 14:7846–7853, 2012.
- ²⁴ L. Shulenburger and T. R. Mattsson. *Phys. Rev. B*, 88:245117, 2013.

- ²⁵ S. J. Cox, M. D. Towler, D. Alfè, and A. Michaelides. *J. Chem. Phys.*, 140:174703, 2014.
- ²⁶ D. Quigley, D. Alfè, and B. Slater. *J. Chem. Phys.*, 141:161102, 2014.
- ²⁷ L. Shulenburger, M. P. Desjarlais, and T. R. Mattsson. *Phys. Rev. B*, 90:140104(R), 2014.
- ²⁸ E. Mostaani, N. D. Drummond, and V. I. Fal'ko. *Phys. Rev. Lett.*, 115:115501, 2015.
- ²⁹ L. Horvathova, M. Dubecky, L. Mitas, and I. Stich. *J. Comp. Theor. Chem.*, 9:390, 2013.
- ³⁰ P. Jurečka, J. Šponer, J. Černý, and P. Hobza. *Phys. Chem. Chem. Phys.*, 8(17):1985–1993, 2006.
- ³¹ T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill. *J. Chem. Phys.*, 132:144104, 2010.
- ³² M. Dubecký. *Acta Phys. Slovaca*, 64:501–574, 2014.
- ³³ J. Xu, M. J. Deible, K. A. Peterson, and K. D. Jordan. *J. Chem. Theory Comput.*, 9:2170–2178, 2013.
- ³⁴ J. Řezáč and P. Hobza. *J. Chem. Theory Comput.*, 9:2151–2155, 2013.
- ³⁵ M. Burkatzki, C. Filippi, and M. Dolg. *J. Chem. Phys.*, 126:234105, 2007.
- ³⁶ M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery. *J. Comput. Chem.*, 14:1347–1363, 1993.
- ³⁷ J. W. Moskowitz and K. E. Schmidt. *J. Chem. Phys.*, 97:3382–3385, 1992.
- ³⁸ C. J. Umrigar and Claudia Filippi. *Phys. Rev. Lett.*, 94:150201, 2005.
- ³⁹ M. Casula. *Phys. Rev. B*, 74:161102(R), 2006.
- ⁴⁰ Lucas K. Wagner, Michal Bajdich, and Lubos Mitas. *J. Comput. Phys.*, 228:3390–3404, 2009. <http://www.qwalk.org/>.